IBM Data Science Capstone Project (Week 4)

Battle of the Neighborhoods (Part 1)

Introduction: My wife and I are currently living in Long Beach, CA and she has a job opportunity in Atlanta, GA. Before we just look for places to live, I would to do some analysis about Atlanta, regarding which neighborhoods are safer.

Problem: I want to see if there is a correlation between top venue categories and the number of crimes committed. For example, if there's a neighborhood that has a higher number of bars/restaurants, will there be a higher crime rate at that particular neighborhood? I will use K-Means algorithm to cluster certain neighborhoods together and compare similar characteristics that each neighborhood obtains to help make a decision. I will also do some Exploratory Data Analysis to identify any trends of crimes, venues or anything else that comes to mind.

Data: The data I will gather to analyze for this project are:

1. Crime data appertaining to Atlanta from January 2018 to December 2018.
2. Geographical data of neighborhoods from Foursquare.

Here are the links of the data I am using:

- https://en.wikipedia.org/wiki/Neighborhood_planning_unit
- http://opendata.atlantapd.org/Crimedata/Default.aspx

With those two links, I will scrape data from those websites so I can build two data frames. I will build a data frame for crime and one for venues which will include latitude and longitude.